UCRL-TR-228283

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

A Systematic Procedure for Assigning Uncertainties to Data Evaluations

W. Younes

February 23, 2007

Disclaimer

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

A systematic procedure for assigning uncertainties to data evaluations

W. Younes
Lawrence Livermore National Laboratory, Livermore, CA 94551
(Dated: 21st February 2007)

In this report, an algorithm that automatically constructs an uncertainty band around any evaluation curve is described. Given an evaluation curve and a corresponding set of experimental data points with x and y error bars, the algorithm expands a symmetric region around the evaluation curve until 68.3% of a set of points, randomly sampled from the experimental data, fall within the region. For a given evaluation curve, the region expanded in this way represents, by definition, a one-standard-deviation interval about the evaluation that accounts for the experimental data. The algorithm is tested against several benchmarks, and is shown to be well-behaved, even when there are large gaps in the available experimental data. The performance of the algorithm is assessed quantitatively using the tools of statistical-inference theory.

I. INTRODUCTION

In this report, we examine the problem of assigning meaningful uncertainties to a data evaluation. If the evaluation can be expressed as a function of a set of parameters, there exist well-established procedures for determining the variability of the evaluation with respect to the parameters. For example, if the evaluation depends linearly on the parameters, the full machinery of linear-regression analysis can be readily applied. In practical applications, however, it is not always possible to evaluate a set of data with a well-defined parametric function. The evaluation may depend on quantities that cannot be easily varied in a continuous manner. For example, the quantum numbers of the states in a discrete level scheme may not be known. Furthermore, it may be necessary to probe the sensitivity of the evaluation model with respect to missing or extraneous levels, without a clear prescription for adding or removing those levels. The evaluation may also be uncertain because it relies on a simplistic picture of the underlying physics. In this case, it is impossible to vary an abstract concept, such as the "degree of realism" of the model, and observe the resulting effect on the evaluation. For those cases where variations of the evaluation are not possible, or not practical, a different approach is needed to quantify the uncertainty in the evaluated results.

As a concrete example, consider the ^{75}As(n,2n) reaction cross-section evaluation [1], show in Fig. 1. The individual data points are taken from several (and in principle independent) measurements spanning four decades. The solid line was obtained from a Hauser-Feshbach calculation using the code STAPRE. For the reasons cited above, it was not practical to attempt to run STAPRE repeatedly, while varying some aspects of the physics it implements, in order to obtain a spread of curves representing the uncertainty in the model calculation. A more useful approach is to take the solid curve as a given, and find the uncertainty band around it, that is most representative of the data points. Recalling the definition of an error bar, a simple prescription for finding the uncertainty band presents itself. In the case of a Gaussian probability distribution, which we will assume is an appropriate distribution in this case, a one-standard-deviation (σ) interval around a mean value $\bar{x}$ (in this case, the evaluation) will contain a fraction of the data equal to

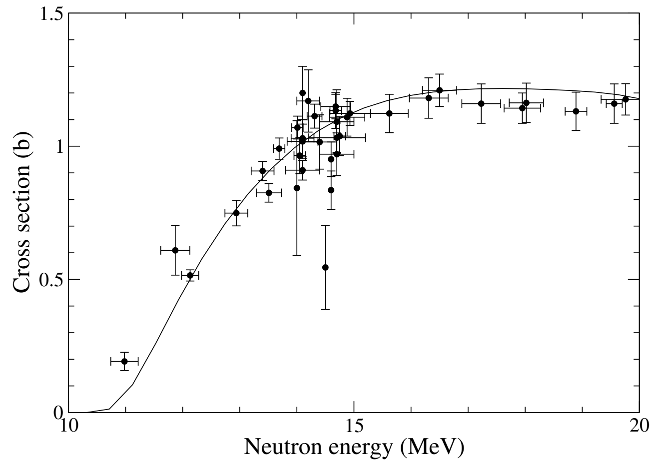

Figure 1: Evaluation of the ^{75}As(n,2n) cross section, taken from [1].

$$\int_{\bar{x}-\sigma}^{\bar{x}+\sigma} dx\, \frac{e^{-(x-\bar{x})^2/\left(2\sigma^2\right)}}{\sqrt{2\pi}\sigma} \approx 0.682689$$

Thus, an uncertainty band for the evaluation in Fig. 1 can be found by "expanding" a region around the solid curve until 68.3% of the data represented by the points in the figure have been included. The main point which will be addressed in the remainder of this report, is how to define and expand an appropriate region around the evaluation curve.

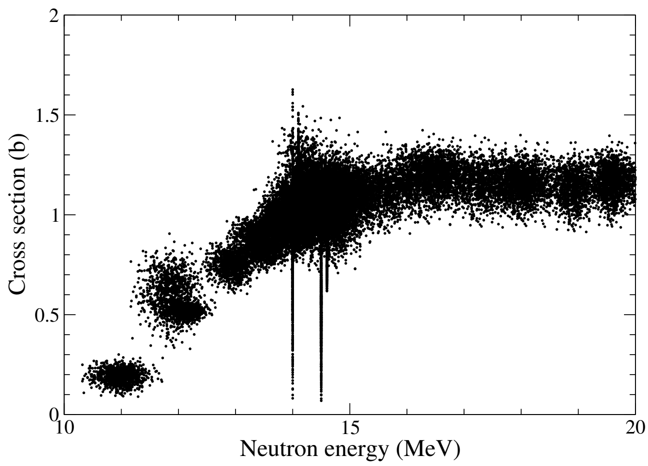

Figure 2: Sampled points generated by a Monte-Carlo procedure from the data in Fig. 1.

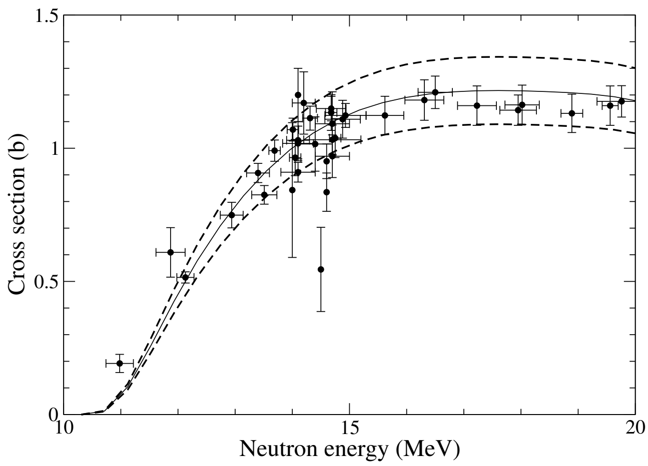

Figure 3: Uncertainty band deduced in [1]. The solid line gives the evaluation, the dashed lines represent the uncertainty band obtained by assuming a global relative uncertainty.

II. METHOD

For the STAPRE calculation and data points in Fig. 1, and uncertainty band was generated by assuming a constant relative uncertainty for all neutron energies [1]. Thus, if the cross-section curve in the figure is given by the function $\bar{y}(E_n)$, where E_n is the neutron energy, the upper (y_+) and lower (y_-) curves delimiting the uncertainty band are given by

$$\begin{aligned} y_+(E_n) &= \bar{y}(E_n)(1+k) \\ y_-(E_n) &= \bar{y}(E_n)(1-k) \end{aligned}$$

where k is a constant. In practice, the value of k was determined in two steps. First, each data point and its associated x and y error bars in Fig. 1 was interpreted as the mean and standard deviation of a Gaussian distribution, and a "cloud" of 1000 points, sampled from this distribution using a Monte-Carlo procedure was generated for each original data point, as shown in Fig. 2. Then, the value of k was increased gradually from 0 until 68.3% of the Monte-Carlo points were included between $y_-(E_n)$ and $y_+(E_n)$. A value of $k = 10.4\%$ was obtained in this way, and the corresponding confidence band is plotted in Fig. 3.

The procedure described above yields, in a relatively straightforward way, a meaningful uncertainty band for the evaluation in Fig. 1. The main limitation of the technique is that it is not very sensitive to the local behavior of the experimental data points. For example, near the reaction threshold ($E_n \approx$10-12 MeV), the first two data points are in poor agreement with the calculated cross-section curve. However, this discrepancy is not reflected in the width of the uncertainty band, as can be seen in Fig. 3. We will now generalize the "global" uncertainty-band procedure described above to take into account local variations in the distribution of experimental data points.

We begin by assuming that, given an evaluation curve $\bar{y}(x)$, the corresponding standard deviation $\sigma(x)$ is a slowly-varying function of the independent variable x. The upper and lower limits of the confidence band are given by

$$\begin{aligned} y_+(x) &= \bar{y}(x) + \sigma(x) \\ y_-(x) &= \bar{y}(x) - \sigma(x) \end{aligned}$$

In order to determine the local value, $\sigma(x)$, of the standard deviation at a given value of x, we generate a cloud of Monte-Carlo points from the experimental data [3], and gradually expand an interval $[x - \Delta x, x + \Delta x]$ centered on x until a statistically-significant number of Monte-Carlo points is contained within the interval. The actual minimum number of points that constitute a statistically significant sample is supplied by the user as an input to the algorithm. Once the interval $[x - \Delta x, x + \Delta x]$ has been constructed, we assume that the interval is small enough that variations in $\sigma(x)$ over the width of the interval can be neglected, and we use a local constant value $\sigma(x) \equiv \sigma$. Now, we can proceed as we did before, slowly increasing the value of σ from zero until 68.3% of the Monte-Carlo points in the local interval $[x - \Delta x, x + \Delta x]$ are included between $y_-(x) = \bar{y}(x) - \sigma$ and $y_+(x) = \bar{y}(x) + \sigma$. This procedure is repeated systematically over the set of discrete points defining the evaluated curve $\bar{y}(x)$, first expanding a local x interval to encompass a statistically-significant Monte-Carlo sample and then expanding the one-sigma y band, and the corresponding uncertainty band for the entire curve is generated [4].

III. BENCHMARKS

In order to test the local-uncertainty-band algorithm described in section II, we construct an evaluation curve $\bar{y}(x)$ and a corresponding standard deviation $\sigma(x)$ using

$$\bar{y}(x) \equiv 100e^{-x/100} \tag{1}$$

$$\sigma(x) \equiv 6 + 5\sin\frac{2\pi x}{200} \tag{2}$$

Note that we could have generated the curves $\bar{y}(x)$ and $\sigma(x)$ in a more consistent manner, starting from a function $y(x) \equiv Ae^{-x/a}$ where A has an expected value $\bar{A}$ and an uncertainty σ_A, and a has an expected value $\bar{a}$ and uncertainty σ_a. The curve $\bar{y}(x)$ is then obtained by setting $A = \bar{A}$ and $a = \bar{a}$ inside $y(x)$, and $\sigma(x)$ corresponds to (independent) variations of the Gaussian random variables A and a in the function $y(x)$. Instead, we have imposed the functional form in Eq. (2) for $\sigma(x)$, because the oscillations in this function should provide a more demanding test for the algorithm. There is, however, a subtle point to make regarding this choice of the form of $\sigma(x)$ which sheds light on the meaning of the uncertainty band generated by the algorithm discussed here, and we will return to this point in section IV. The evaluation in Eq. (1) and its uncertainty band, deduced using Eq. (2), are plotted in Fig. 4. From these curves, Monte-Carlo points are generated as follows: a value of the abscissa x is sampled from a uniform distribution in the interval $[0, 300]$, for that value of x a value of the ordinate y is sampled from a Gaussian distribution with mean $\bar{y}(x)$ and standard deviation $\sigma(x)$. A total of 10^5 Monte-Carlo points was generated in this manner, and the points are plotted in Fig. 5. The test then is to see whether the uncertainty-band algorithm can reconstruct the dashed lines in Fig. 4, starting from the Monte-Carlo points in Fig. 5. For this test, the number of statistically significant points in a local interval was set at 200. The $y_-(x)$ and $y_+(x)$ curves reconstructed by the algorithm are compared in Fig. 6 to the original curves from Fig. 4. In this case, the uncertainty band was reconstructed with astounding accuracy. Small fluctuations of the reconstructed curves about the original lines can be attributed to the Monte-Carlo procedure used to generate the points in Fig. 5.

Next, we test the ability of the algorithm to reconstruct the original uncertainty band when large portions of the data are missing. In Fig. 7, we have reduced the Monte-Carlo points from Fig. 5 by omitting those points whose abscissa falls within the intervals $[25, 75]$, $[125, 175]$, or $[225, 275]$. The uncertainty-band algorithm was run, again requiring at least 200 points in each local interval, and the result is compared in Fig. 8 with the original uncertainty band. Despite the gaps in the Monte-Carlo data, the uncertainty band is very-well reproduced by the algorithm.

Finally, we return to the cross-section evaluation in Fig. 1, and apply the local-uncertainty-band algorithm, with the usual 200-point requirement for the local intervals. The resulting uncertainty band is plotted in Fig. 9, and should be compared with the band in Fig. 3. The uncertainty band in Fig. 9 is a much better estimate of the extent to which the evaluation represents the available data. In particular, we see that in the range $E_n \approx$10-12 MeV, the band becomes larger where the evaluation and data are in poor agreement, as expected. However, for a more formal assessment of the algorithm's performance in this case, the interested reader should consult appendix A. One valid criticism that can be leveled against the uncertainty band in Fig. 9 is that it remains relatively broad at the reaction threshold ($E_n = 10.382$ MeV), where we know that the cross section must go to zero. There are several ways to mitigate this problem. First, a new evaluation that passes closer to the data points with $E_n =$ 10.98 and 11.87 MeV would lead to an uncertainty band that is much narrower near and at threshold. Second, and artificial data point with zero x and y error bars could be added at threshold, forcing the uncertainty band to shrink to zero width at that point. Third, near threshold, the assumption of a slowly varying $\sigma(x)$ may no longer be valid. In that case, it may be more appropriate to assume a constant relative uncertainty near threshold, in such a way that the deduced $\sigma(x)$ transitions smoothly from the constant relative uncertainty to the constant absolute uncertainty model as a function of x.

In addition to producing uncertainty bands, the algorithm described in section II can also be used to construct confidence intervals, which can be thought of in this context as a generalization of the concept of an error bar. For example, in Fig. 10, the 95% confidence interval is plotted, and represents a band containing 95% of the Monte-Carlo points. In this particular case, the 95% confidence band looks more or less like a wider version of the (68.3%) uncertainty band, except near $E_n = 14$ MeV, where a large spread in the data causes the band to jut out noticeably.

IV. DISCUSSION AND CONCLUSION

We noted in section III that the uncertainty band for the fictitious evaluated curve in Fig. 4 could have been obtained by varying the parameters in a given model. This approach can always be used to construct the uncertainty band, given a parametric model, and data from which the variability of the parameters can be deduced. In fact, if we assumed the exponential model $y(x) \equiv Ae^{-x/a}$ with Gaussian distributions for the parameters A and a, we could never reproduce the "data" points in Fig. 5. We would then be led to conclude that the model, the data, or both are not correct. Conversely,

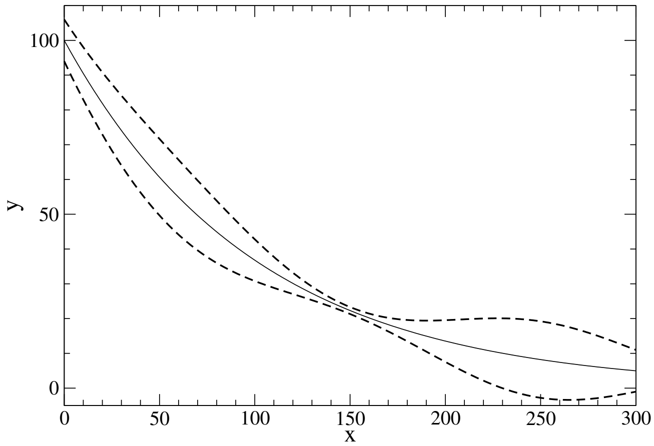

Figure 4: Functional forms used to test the confidence-band algorithm. The solid line represents $\bar{y}(x)$, given by Eq. (1), and the dashed lines represent the curves $y_-(x)$ and $y_+(x)$ given by Eq. (2).

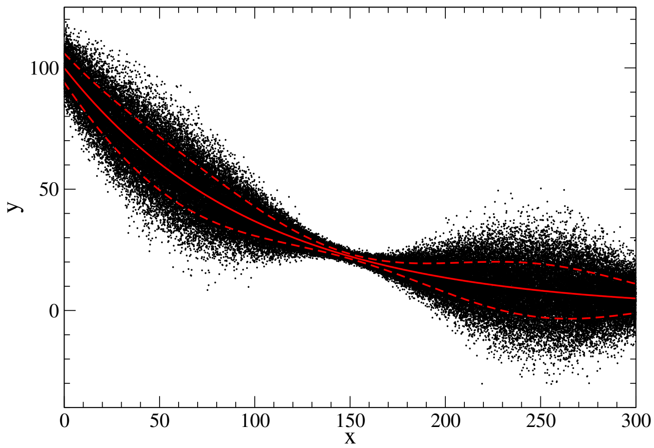

Figure 5: Data points generated by a Monte-Carlo procedure using the functional forms plotted in Fig. 4. The curves from Fig. 4 are overlayed on top of the data for reference.

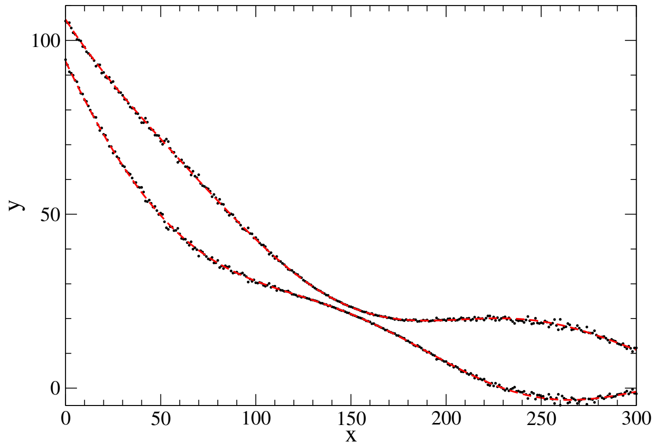

Figure 6: Reconstruction of the uncertainty band from the data in Fig. 5, using the confidence-band algorithm. The dashed lines represent the original uncertainty band, and the discrete points represent their reconstruction.

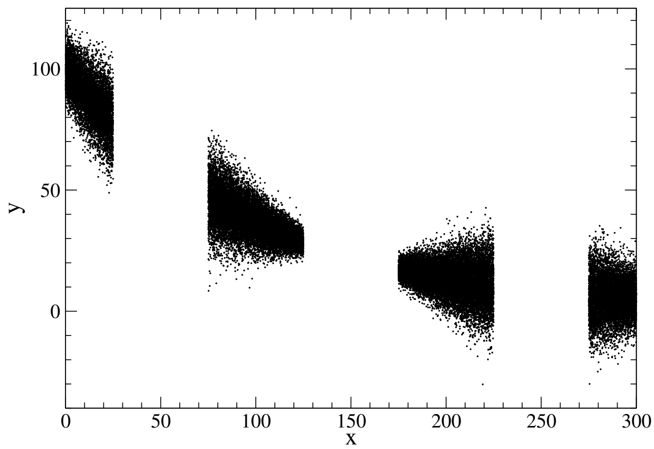

Figure 7: New data set, generated from the data in Fig. 5, by removing significant ranges of points. This new data set is used to test the behavior of the confidence-band algorithm when faced with missing data.

the algorithm described in this report cannot be used to validate either the evaluation curve or the data. Instead, the algorithm constructs an uncertainty band, based on a very different premise: the method assumes that the model is not known but that the standard deviation is a slowly-varying function of the independent variable x, and attempts to generate uncertainties that are consistent with the available data.

We have developed and implemented an algorithm that constructs a meaningful uncertainty band, given an evaluation curve and a set of experimental data points with x- and y-error bars. The algorithm is sensitive to local variations in the distribution of the experimental data around the evaluation curve. The algorithm has been tested with several benchmarks. First, it has been shown that the algorithm can correctly reconstruct the uncertainty band from a parent distribution, given a Monte-Carlo sample from that distribution. Next, the algorithm was shown to closely reproduce the uncertainty band,

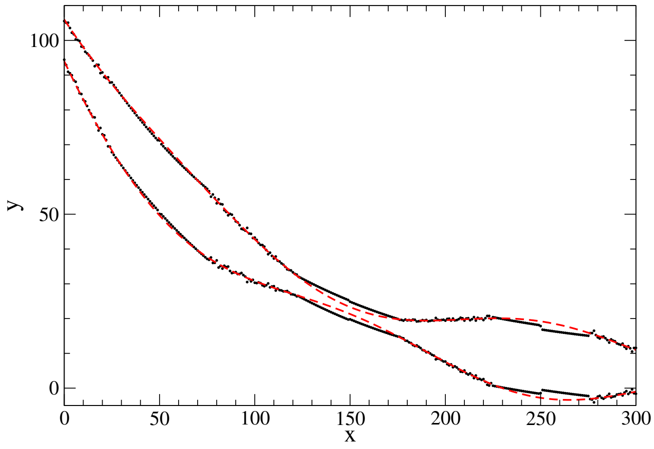

Figure 8: Uncertainty band reconstructed from the reduced data in Fig. 7. The dashed lines represent the original uncertainty band, and the discrete points represent their reconstruction.

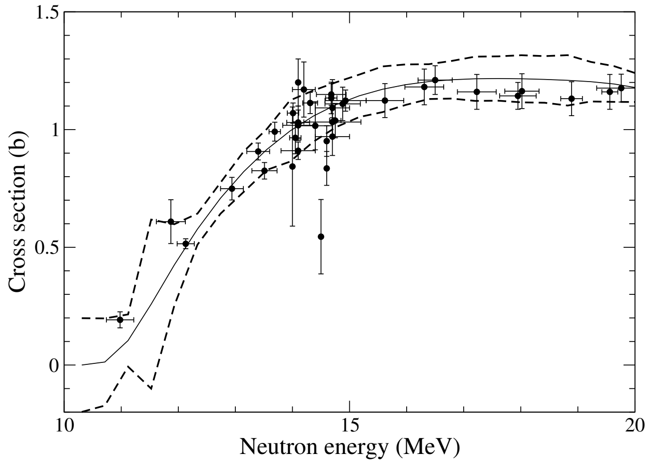

Figure 9: Uncertainty band (dashed lines) reconstructed from the data in Fig. 2. The evaluation curve (solid line) and original data (solid points) from Fig. 3 are shown for reference.

even when sizable portions of the Monte-Carlo sample are discarded. Finally, the method was applied in a realistic situation by providing an uncertainty band for the ^{75}As(n,2n) cross-section, evaluated in [1]. The algorithm described here should be used when the evaluation cannot be represented by a parametric model whose parameters can be meaningfully or practically varied to reproduce the spread in experimental data.

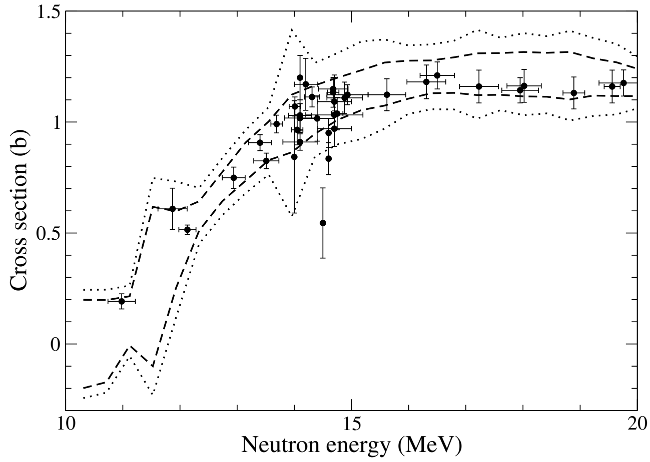

Figure 10: Plot of the 95% confidence interval (dotted lines) around the evaluation curve, shown along with the uncertainty band (dashed lines) from Fig. 9.

Appendix A: STATISTICAL INFERENCE ON THE PERFORMANCE OF THE ALGORITHM

In this appendix, we formulate a hypothesis test for the performance of the band-construction algorithm (see, e.g., [2] for a detailed description of the methodology). We take as a criterion for success the extent to which the fraction of all Monte-Carlo points included within the band approaches the theoretical limit of $p_0 \approx 0.682689$. In general, the algorithm can fail if the number of Monte-Carlo points is too small, if the minimum number of points in a local interval specified by the user is too small, or if the discrete points about which the local intervals are expanded are too far apart (leading to the “kinks” in the band in Fig. 9).

To perform the test, the upper- and lower-bound curves $y_-(x)$ and $y_+(x)$ produced by the algorithm are splined, and the overall proportion p of the N Monte-Carlo points that fall between the two curves is extracted. For the results plotted in Fig. 9, we find $N = 37000$ and $p = 0.6429$. With this information, we can test the null hypothesis that $p = p_0$, against the alternative $p \neq p_0$ at, for example, the $\alpha = 0.01$ level of significance (i.e., the probability of incorrectly rejecting the null hypothesis if it is true is $\alpha = 0.01$). We form the random variable

$$z = \frac{Np - Np_0}{\sqrt{Np_0\,(1 - p_0)}} \tag{A1}$$

which, by the central-limit theorem has a Gaussian probability distribution with mean 0 and standard deviation 1. At the chosen level of significance, the null hypothesis must be rejected if the value of z in Eq. (A1) is lower than -2.575 or greater than +2.575 (which defines the so-called

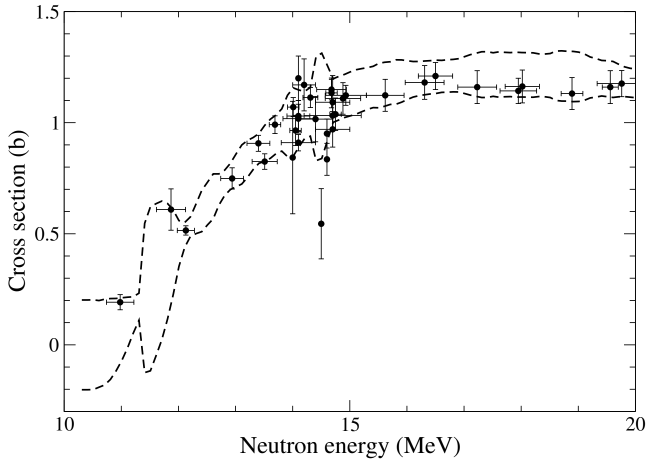

Figure 11: Improved estimate of the uncertainty band (see text). This result should be compared with the one displayed in Fig. 9.

critical region for a Gaussian distribution with $\alpha = 0.01$). Using the values $N = 37000$ and $p = 0.6429$ in Eq. (A1), we find $z = -16.4441$, which clearly falls within the critical region. Therefore, we must reject the null hypothesis that $p = p_0$ at the 0.01 level of significance, and conclude that the uncertainty in Fig. 9 should be re-calculated with a larger Monte-Carlo sample and a finer step size for the evaluation curve.

In light of this conclusion, we increase the number of Monte-Carlo points generated for each experimental data point from 1000 to 2000, and we spline the evaluation curve and refine the step size along the x axis to 0.1 MeV. The new uncertainty band is shown in Fig. 11, and is qualitatively similar to the one shown in Fig. 9. However, for this new band, we extract an overall proportion $p = 0.682309$ of the $N = 74000$ Monte-Carlo points within the uncertainty band. Using Eq. (A1), a value of $z = -0.222098$ is found in this case, which is well outside the critical region, and we therefore accept the null hypothesis. In other words, we can say that, at the 0.01 level of significance, the uncertainty band in Fig. 11 includes the known experimental data 68.3% of the time, as it should.

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

[1] W. Younes, P. E. Garrett, J. A. Becker, W. E. Ormand, F. S. Dietrich, R. O. Nelson, M. Devlin, and N. Fotiades, "The ^{75}As(n,2n) cross sections into the ^{74}As isomer and ground state", LLNL Tech. Rep. no. UCRL-ID-154061 (2003).

[2] R. E. Walpole, R. H. Myers, Probability and Statistics for Engineers and Scientists, third Ed., chapter 8, Macmillan Publishing Company, New York (1985).

[3] The procedure used to generate the Monte-Carlo samples from a given set of experimental data points with both x- and y-error bars was implemented in the C++ code **mcdata**.

[4] The uncertainty-band algorithm was implemented in the C++ code **errband**.